

## SEEDS Critical Capabilities List

June 17, 2002

### Participant notes grouped (not reduced) and organized by **key concepts**

- **Faster access** to the data by the user, delivery
- Online access to information (not data)
- Near real-time access to data, immediate access to relevant data
- Being able to access manageable amounts of data
  - Data volumes are large
  - Need methods to provide manageable amounts of data to fulfill any user's need
- Ability to obtain data & information (and knowledge about the data & info) quicker
  - This may require use of AI techniques (expert systems) to generate quality products (QA with %certainty)
  - The "best guess" products are generated automatically and include real-time load of collaborative inputs, adjustments of parameters (coefficients, processing decision trees, etc) and statistical reference of quality (a statement to the % certainty of the product).
  - Extend capability to international community
- Enable **searches** over multiple archives
- Content-based search & retrieval, event based search
- Interface support for automated **query** & retrieval of data
- Better (seamless) interaction with/connectivity to the international community
  - Eg, common metadata, search, access, etc. in both directions
  - Data in / Truth out
- Data **discovery** – expert knowledge guide user to appropriate data sets
  - Relevancy rankings on search results – expert guidance, better cataloging of distributed info systems
  - Easily resolve/filter missing data
  - Reduce search result set
- Closer/**seamless integration** between catalog systems & data access
  - Catalog + access = seamless
- Recognize events without reprocessing, without computing new metadata
  
- **Ease of use**
- Make the **data adapt** to regularly available packages
  - Rather than waiting for commercial products (eg IDL) to adapt to data
- Ability to filter data to information to knowledge
  - NASA will have too much data
  - Too much data provides no information
- **Innovative software tools**
  - Package & provide tools with data that make data accessible & usable, as easy to use as the WWW
  - Easy to use for broad user population

## SEEDS Critical Capabilities List

June 17, 2002

- Able to access & combine data from variety of sources: GIS, remote sensed, etc
  - Low cost & easy to find/install
  - Better support / adaptation to commercial packages (eg HDF5)
- Novice users – easy access
- Interactive performance
- Voice recognition techniques
- Easy data fusion
- Earth science information science curriculum
- More/better **Web-based services** for data
  - Smart service chaining
  - Integrated, web-based, data services
  - Data + tools + easy user ala WWW
  - Extend to international community
- Capability to generate a user community specific portal to a subset of data and services needed
  - Interfaces & toolkits to generate portal view from the comprehensive systems
- **Semantic web** for scientific data
- More semantic information
  - Understanding data format, structure, content, science parameters, etc. (maybe dynamic links between data & documents?)
  - Mapping/cross referencing of vocabularies
- **Transparent format & location**
- Data format transparency
  - Very high level data objects – easy for end users to work w/ data without having to know data formats – domain semantics
- Standardize data format
- Data **format**
  - Fewer standards
  - <More> translators
- Develop & organize translators for everything
  - Between data formats
  - Between various interfaces
  - Ability to distribute in any format
- Seamless data usage
  - Being able to pull datasets together, regardless of origin
- Agreement on archive formats / **data model**
  - Data models to support emerging data collection methods (non-satellite, nano tech, field tech, non-gridded), like smart seeds, etc, compatible with data models for other types of data
- Data **verification**
- Access to verified / validated data services (trusted)
- Quality assessment

## SEEDS Critical Capabilities List

June 17, 2002

- Access and validation services – quality labels
- Alternatives to Firewalls
  - Access to secure/classified data
- **On-demand, custom products** generated by user request
  - Products can be combined products
  - Extracted/mined, combined form multiple sources
  - System to specify product needed (plain English)
  - Personalized data services (fusion and processing)
- On-demand re-projection
- No data system at all (bypass, direct broadcast to anyone, on-satellite DAACs)
  - Data purely on satellites, direct broadcast to user, apps, archives
  - “Click here to create the data processing system” – toolkit to create data product a zero cost
  - Tools to avoid/detect coding errors – high level manipulation, language
- Must be adaptable to ever-**evolving technology**
  - Two years ago XML was not available, yet today it is the way of interfacing; yet ECS does not use XML inn the mainstream system
- Ability for tools to evolve more easily
- Standards & protocols
- Flexible
- Easily upgradeable
- Simplicity of design (avoid gratuitous complexity)
- **High performance** data processing **capabilities**
  - Emerging parallel architectures, clusters, grids
- Desktop storage
- Computing power (processing speed)
- Input/output
- Better **distribution** than mail
- Store and compress large objects – subset decompression
- **Network**
  - Broadband
  - Wireless
- Network bandwidth 10x
- Electronic transfer 10x
- Ability to subscribe / push data to wherever one is
  - **Wireless** ingest of data/info to palm pilot via subscription services for anywhere in the world one happens to be; data/info include any phenomena of interest in any part of the world

SEEDS Critical Capabilities List  
June 17, 2002

- **Always up**
- Stability
- Insure survive-ability of data / metadata

## SEEDS Critical Capabilities List

June 17, 2002

### Raw list from participant comments during workshop session

1. Online access to information (not data)
2. Novice users – easy access
3. More/better web-based services
4. Easy data fusion
5. Easily resolve/filter missing data
6. Better human/machine interfaces – interactive/immersive
7. Relevancy rankings on search results – expert guidance, better cataloging of distributed info systems
8. Automated operations
9. Smart service chaining
10. Data + tools = easy use (ala WWW)
11. Standardize data format
12. Better distribution than mail
13. Catalog + access =>seamless
14. Alternative to Firewalls
15. Access to secure/classified data
16. Immediate access to relevant data
17. Ability to distribute in any format
18. Bypass data systems – direct broadcast to anyone - toolkits for building data systems (on-satellite DAACs)
19. Reduce data and info to knowledge / AI for QA %certainty
20. Ability to subscribe push data to wherever one is
21. High perf processing capabilities on emerging tech.
22. Data models to support emerging data collection methods – non-satellite, nano tech field tech, non-gridded
23. Custom products, extracted (mined), combined from multiple sources, also system to specify what needed (plain English)
24. Recognize events w/o reprocessing – w/o computing new metadata
25. 4, 3, 15 Extend to international community
26. Restate #19 – reducing search result set
27. Better support / adaptation to commercial packages (e.g. HDF5) how to take advantage of commercial packages w/o having to wait for them to adapt to format
28. Personalized data services (fusion and process)
29. Access and validate services – quality labels
30. Always up
31. Semantic web
32. Content based search and retrieval w/ voice
33. Store and compress large objects – subset decompression
34. Very high level data objects – easy for end users to work w/ data without having to know data formats – domain semantics
35. Tools to avoid/detect coding errors – high level of manipulation language
36. Simplicity of design (avoid gratuitous complexity)

SEEDS Critical Capabilities List  
June 17, 2002

Raw notes from participant handouts returned at workshop

- Integrated, web-based, data services
- Capability to generate a user community specific portal to a subset of data and services needed
  - Interfaces & toolkits to generate portal view from the comprehensive systems
- On-demand, custom products generated by user request
  - Products can be combined products
- Must be adaptable to ever-evolving technology
  - Two years ago XML was not available, yet today it is the way of interfacing; yet ECS does not use XML in the mainstream system
- Make the data adapt to regularly available packages
  - Rather than waiting for commercial products (eg IDL) to adapt to data
- Ability to filter data to information to knowledge
  - NASA will have too much data
  - “Ugly Duckling Theorem” (too much data provides no information)
- Near real-time access to data
- Faster access of the data to the user, delivery
- Closer/seamless integration between catalog systems & data access
- More semantic information
  - Understanding data format, structure, content, science parameters, etc. (maybe dynamic links between data & documents?)
  - Mapping/cross referencing of vocabularies
- Transparent format & location
- Being able to access manageable amounts of data
  - Data volumes are large
  - Need methods to provide manageable amounts of data to fulfill any user's need
- Seamless data usage
  - Being able to pull datasets together, regardless of origin
- Develop & organize translators for everything
  - Between data formats
  - Between various interfaces
  - Etc
- Package & provide tools with data that make data accessible & usable, as easy to use as the WWW
- Data format
  - Fewer standards

## SEEDS Critical Capabilities List

June 17, 2002

- <More> translators
- Interface support for automated query & retrieval of data
- Better (seamless) interaction with/connectivity to the international community
  - Eg, common metadata, search, access, etc. in both directions
  - Data in / Truth out
- Tools
  - Easy to use for broad user population
  - Able to access & combine data form variety of sources: GIS, remote sensed, etc
  - Low cost & easy to find/install
- High performance data processing capabilities
  - Emerging parallel architectures, clusters, grids
- Data models to support emerging data collection methods, like smart seeds, etc, compatible with dat models for other types of data
- Earth science information science curriculum
- Ability for tools to evolve more easily
- Ability to obtain data & information (and knowledge about the data & info) quicker
  - This may require use of AI techniques (expert systems) to generate quality products
  - The “best guess” products are generated automatically and include real-time load of collaborative inputs, adjustments of parameters (coefficients, processing decision trees, etc) and statistical reference of quality (a statement to the % certainty of the product).
- Wireless ingest of data/info to palm pilot via subscription services for anywhere in the world one happens to be; data/info include any phenomena of interest in any part of the world
- Network
  - Broadband
  - Wireless
- Standards & protocols
- Computing power (processing speed)
- Input/output
- Desktop storage
- Data fusion
- Innovative software tools
- Content-based search & retrieval, event based search
- Voice recognition techniques
- Semantic web for scientific data
- Always up
- Stability

SEEDS Critical Capabilities List  
June 17, 2002

- Network bandwidth 10x
- Electronic transfer 10x
- Ease of use
- Interactive performance
- Data verification
- Flexible
- Easily upgradeable
  
- Quality assessment
- No data system at all
  - Data purely on satellites, direct broadcast to user, apps, archives
  - “Click here to create the data processing system” – toolkit to create data product a zero cost
- Insure survive-ability of data / metadata
- Web0based services on data
- Agreement on archive formats / data model
- Enable searches over multiple archives
  
- Data format transparency
- On-demand re-projection
- Access to verified / validated data services (trusted)
- Data discovery – expert knowledge guide user to appropriate data sets